PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 5 : H04N 9/79, 5/76

A1

(11) International Publication Number: WO 94/11995

(43) International Publication Date: 26 May 1994 (26.05.94)

(21) International Application Number: PCT/US93/11267

(22) International Filing Date: 17 November 1993 (17.11.93)

(30) Priority data:
07/977,885 17 November 1992 (17.11.92) US

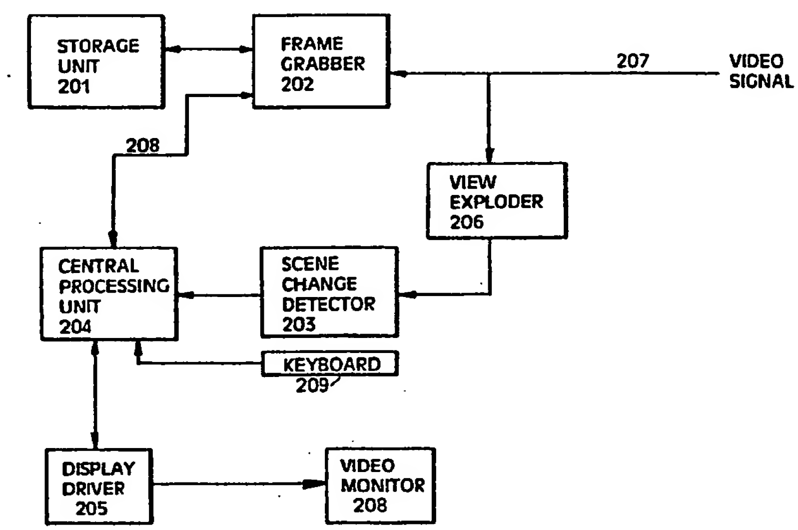
(71)(72) Applicants and Inventors: DUBNER, Harvey [US/US]; 449 Beverly Road, Ridgewood, NJ 07450 (US). DUBNER, Robert [US/US]; 13 Westervelt Place, Westwood, NJ 07675 (US).

(74) Agent: DREYER, Douglas, P.; Gottlieb, Rackman & Reisman, 1430 Broadway, New York, NY 10018 (US).

(81) Designated States: European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

**Published**
*With international search report.*
*Before the expiration of the time limit for amending the claims and to be republished in the event of the receipt of amendments.*

(54) Title: VIDEO LOGGING SYSTEM AND METHOD THEREOF




(57) Abstract

An improved video tape logging device (2) comprising a recording member (102) for recording audio information and video information (207). The audio information and video information (207) are linked together, so that the videotape (102) can be logged by either audio cut points or video cut points. The device also includes a videotape scene change detector (203), coupled to a memory device (201) for storing a plurality of selective frames, preferably one from each scene. Also included in the system is circuitry (205) for displaying a plurality of frames simultaneously on a video display monitor (208) or computer (101), which frames have been reduced from the complete frames by selective sampling. The generation of a video signal representative of the audio information is provided. The video signal clearly delineates the ends and beginnings of words to facilitate finding cut points on a videotape. Annotation of stored frames is also provided.

*FOR THE PURPOSES OF INFORMATION ONLY*

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

| | | | | | |
|---|---|---|---|---|---|
| AT | Austria | GB | United Kingdom | MR | Mauritania |
| AU | Australia | GE | Georgia | MW | Malawi |
| BB | Barbados | GN | Guinea | NE | Niger |
| BE | Belgium | GR | Greece | NL | Netherlands |
| BF | Burkina Faso | HU | Hungary | NO | Norway |
| BG | Bulgaria | IE | Ireland | NZ | New Zealand |
| BJ | Benin | IT | Italy | PL | Poland |
| BR | Brazil | JP | Japan | PT | Portugal |
| BY | Belarus | KE | Kenya | RO | Romania |
| CA | Canada | KG | Kyrgystan | RU | Russian Federation |
| CF | Central African Republic | KP | Democratic People's Republic of Korea | SD | Sudan |
| CG | Congo | | | SE | Sweden |
| CH | Switzerland | KR | Republic of Korea | SI | Slovenia |
| CI | Côte d'Ivoire | KZ | Kazakhstan | SK | Slovakia |
| CM | Cameroon | LI | Liechtenstein | SN | Senegal |
| CN | China | LK | Sri Lanka | TD | Chad |
| CS | Czechoslovakia | LU | Luxembourg | TG | Togo |
| CZ | Czech Republic | LV | Latvia | TJ | Tajikistan |
| DE | Germany | MC | Monaco | TT | Trinidad and Tobago |
| DK | Denmark | MD | Republic of Moldova | UA | Ukraine |
| ES | Spain | MG | Madagascar | US | United States of America |
| FI | Finland | ML | Mali | UZ | Uzbekistan |
| FR | France | MN | Mongolia | VN | Viet Nam |
| GA | Gabon | | | | |

# VIDEO LOGGING SYSTEM AND METHOD THEREOF

## BACKGROUND OF THE INVENTION

The present invention relates to a scene change detector for audio/video tapes, and more particularly, to an improved method and apparatus for detecting scene changes and assisting an operator in logging the scenes of a videotape program.

In preparing a television program, movie, documentary or commercial from raw videotape, it is typically required to log many hours of prerecorded scenes from one or more videotapes. Considerable time must be spent in locating the start and end of each scene, and once that is accomplished, in identifying the contents of the scene.

Attempts have been made in the past to automate the process of scene change detection. One prior art arrangement is shown in U.S. Patent No. 4,920,423 to Shiota issued on April 24, 1990. The Shiota arrangement describes a system which compares adjacent frames in a long videotape to each other. When there is enough of a change between adjacent frames, the Shiota arrangement concludes that a scene change has occurred and thereupon records the first frame of the new scene. After the first frame from numerous scenes has been recorded, a printout is made of a plurality of frames, one from each scene.

The Shiota arrangement assists in automating the complex task of video logging by allowing an operator to examine a print of one frame selected from a plurality of scenes, so that the operator may determine the sequence of scenes on the videotape, and further, may decide which scenes to edit out, to reorder, to not use, etc.

While the Shiota system is a significant step in the right direction, there are still several drawbacks

the right direction, there are still several drawbacks in such a system. First, scene changes are detected by correlating individual pixels of adjacent frames. If the scene changes only slightly, only a relatively small number of pixels will change and the correlator will not properly detect that a scene change has actually occurred.

Another prior arrangement directed to this problem is U.S. Patent No. 4,698,664 issued to Nichols et al. on October 6, 1987. This patent describes an arrangement whereby a sequence of frames is recorded and displayed on a video screen. However, there are several drawbacks to the Nichols arrangement as well. First, Nichols records every frame rather than just recording frames at scene changes or predetermined intervals. Thus, while the Nichols system has made strides in the right direction, it is a commercially undesirable system, in part because in grabbing every frame and storing every frame the computer system used would require too much storage for a long videotape, and would display too much unrequired "extra frames." Accordingly, the computer speed would be compromised and the entire process would remain quite time consuming for the operator to edit and rearrange the videotape.

Furthermore, none of the prior art devices provide a device that logs scenes based on the audio signal. For example, in an interview situation where a camera is stationary and the video scene does not change, the prior art devices do not provide any help in detection or logging of the videotape. The present invention allows the user to hear the sound associated with the video in order to help in logging of the videotape. The system further includes a graph of the audio sound, such that the exact frame on which a word ends may be logged.

Accordingly, there exists a need for an improved and convenient videotape logging system which will detect scene changes, record a relatively small amount of information to be used by the operator, and yet still allow the operator to conveniently and easily edit and rearrange the scenes on the videotape, and that is capable of using the audio signal to help log the videotape.

## SUMMARY OF THE INVENTION

Generally speaking, in accordance with the instant invention, an improved audio/video scene change detection and logging system is provided. An audio/video scene change detection and logging apparatus for logging particular audio and video information from an audio/video source includes a member for recording selected video information that is received from the audio/video source. A member is provided for recording the audio information from the audio/video source and connecting the audio information with the video information. A member is also provided for producing a waveform representing the recorded audio information.

In accordance with the invention, scene changes are detected and at least one frame from each scene is stored electronically, for example, in a personal computer memory. The plurality of frames, one from each scene, are then reduced in size and displayed, several at a time, on a video monitor.

In accordance with an optional enhancement, scene changes are more accurately detected by analyzing only a portion of the frame rather than the entire frame. By not analyzing the entire frame in order to detect scene changes, the scene change detector can be made more sensitive.

In another optional enhancement, frames are re-

corded at preset time intervals rather than only at scene changes. Frames are then reduced in size so that a plurality of such frames can be displayed on a video monitor for logging. Such an arrangement provides the operator with an accurate depiction of the progression of the scenes along the videotape. As these frames are displayed, scene change information can be shown, as well as frame numbers, or other relevant data.

In another embodiment, audio information from the videotape is also digitized and stored in a personal computer. Other enhancements are optionally provided for efficient storage, annotation, and display of the digitized and recorded frames.

Accordingly, it is an object of the present invention to provide a highly accurate scene changer detection system that allows the user to listen to the audio as well as seeing unitary frames of video grabbed at predetermined intervals.

It is another object of the invention to provide a detection and logging system that can be used with even relatively slow computer systems (such as a computer containing a 386 microprocessor).

Yet another object of the invention is to provide such a system that is relatively inexpensive to manufacture.

Still other objects and advantages of the inventions will in part be obvious and will in part be apparent from the specification and drawings.

The invention accordingly comprises the features of construction, combination of elements, and arrangement of parts which will be exemplified in the construction hereinafter set forth, and the scope of the invention will be indicated in the claims.

**BRIEF DESCRIPTION OF THE DRAWINGS**

For a fuller understanding of the invention, reference is had to the following description taken in connection with the accompanying drawings, in which:

FIG. 1 shows a block diagram of an illustrative embodiment of the present invention;

FIG. 2 shows a more detailed conceptual block diagram of the major components of the present invention;

FIG. 3 shows a conceptual view of a plurality of pixels being displayed on a video monitor, wherein said pixels form the viewable image;

FIG. 4 shows a more detailed engineering diagram of a scene change detector;

FIG. 5 shows a monitor displaying a plurality of "postage stamp" images in accordance with the present invention;

FIG. 6 depicts a monitor displaying a plurality of postage stamp scenes which are a subset of those shown in FIG. 5, with other optional enhancements illustrated;

FIG. 7 is a flow chart illustrating the steps involved in storing the signal; and

FIG. 8 depicts a monitor displaying a plurality of postage stamp scenes and an audio channel signal in accordance with a further embodiment of the invention.

**DETAILED DESCRIPTION OF THE PREFERRED EMBODIMENT**

FIG. 1 is a system overview illustrating the basic building blocks of a video scene change detection and logging system in accordance with the present invention. The arrangement comprises a personal computer 101 linked for two-way communications with a videotape player (VTP) 102. Personal computer 101 is capable of storing video frames which are transmitted from VTP 102. Optional large screen monitor 103 may

be utilized to display one or more video frames from VTP 102, or to project the entire moving image output from VTP 102.

The invention contemplates a video image received from a videotape which is "normally" an analog source. However, the image could be received from a videotape player in the form of a digital video signal. In either case, the video image is stored digitally. For example, a digital version of every frame is stored electronically, in a computer memory unit such as a hard disk drive, PROM, videodisc, CD ROM, etc. In that case, all of the scene change functions described herein could be accomplished by the software. Alternatively, if the programming functions are burnt into the hardware, the same results can be accomplished at much greater speed.

FIG. 2 is a block diagram illustrating the major operating components of the video scene detection and logging system. The circuit elements carrying out the functional blocks shown in FIG. 2 are preferably mounted on a conventional circuit board installed within personal computer 101 of FIG. 1. Alternatively, a separate "box" may be provided containing all or any portion of the components of FIG. 2, or a custom built computer may be constructed for implementing the functional blocks shown in FIG. 2. The particular implementation is a function of various parameters and is not critical to the present invention.

With particular reference to FIG. 2, the center of the system is the central processing unit (CPU) 204. CPU 204 is essentially the brain of the computer. CPU 204 is electrically coupled to storage unit 201, which is for example a conventional random access memory (RAM). CPU 204 controls storage unit 201 with the appropriate timing and control signals as is well known in the art of digital computer hardware. A

display driver 205 is electrically coupled to CPU 204 and receives instructions from CPU 204 that result in the digital signal that is output to a video monitor 208. Video monitor 208 then displays the image.

CPU 204 is electrically coupled to frame grabber 202, scene change detector 203 and keyboard 209. Scene change detector 203 is electrically coupled to view exploder 206 which receives video signal 207. Frame grabber 202 also receives video signal 207.

FIG. 2 also includes scene change detector 203 for detecting changes in scenes of the video signal 207, frame grabber 202 for extracting a specified frame from a particular scene and transmitting the selected frame to storage unit 201 for later use, and a view exploder 206 described in more detail later herein below.

In operation, a video signal 207 is output for example by VTP 102 of FIG. 1 and is transmitted to frame grabber 202 and view exploder 206. The video signal is preferably of the National Television Standards Committee (NTSC) type but may be varied as desired. View exploder 206 may be selectively turned on or off. View exploder 206 provides an optional enhancement that in essence varies the sensitivity to the scene change detector. For purposes of the present explanation, view exploder 206 is assumed to be in the "off" position. When in the off position, view exploder 206 simply passes video signal 207 to the next block of the system, scene change detector 203, without manipulation of the signal.

Video signal 207 is concurrently transmitted to frame grabber 202 and scene change detector 203. Scene change detector 203 detects each scenes change in video signal 207 substantially instantaneously. Upon such detection, scene change detector 203 signals CPU 204 that a scene change has been detected. CPU

204 then instructs frame grabber 202 via control line 208 that a scene change has just occurred. Frame grabber 202 will then select the next frame from video signal 207 and forward said frame to storage unit 201. In the most preferred embodiment, frame grabber 202 at least temporarily stores each frame grabbed in storage unit 201 and gives the storage unit a time stamp. The time stamp is provided by the internal clock of the computer. CPU 204 is also connected to the internal clock of the computer and upon indication of a scene change from scene change detector 203, CPU 204 marks the first frame of the scene with a scene change designation.

Frame grabber 202 receives a video signal and breaks it down to a single frame. The entire frame may be output or alternatively, a condensed version thereof may be output. For example, if it is desirable to save storage space or to record only a condensed version of the frames, frame grabber 202 could store every $n^{th}$ pixel of the frame being grabbed, where n is a number in the approximate range of 4 or 5. Storing every fifth pixel, for example, will provide sufficient information to reconstruct the frame, although the reconstructed version will lose some resolution when compared with the original frame comprised of all the pixels. Actual storage and compression technology is discussed in more detail below.

Due to the computerized nature of the video scene detection and tape logging system of the present invention, the frames stored in memory can vary greatly. In the most preferred and basic form of the invention, frame grabber 202 grabs one frame at predetermined intervals and forwards it to storage unit 201 where it is stored in digital form. Simultaneously with frame grabbing and storing, the video signal 207 is received by the scene change detector 203 in order to determine

scene change locations.

After storage of a full videotape worth of frames, storage unit 201 will contain many stored frames, each frame being a compressed version of the original frame appearing in video signal 207. Upon operator command, the plurality of frames stored in storage unit 201 may be transmitted through CPU 204 to display driver 205 for display on video monitor 208.

The stored frames may be displayed on a personal computer, a large screen monitor such as that noted in FIG. 1, or by any other suitable means. The operator may then view the frames desired. For example, every scene grabbed may be displayed, or each scene designated as a scene change may be viewed. The ability to view only desired frames assists in logging and reordering the videotape.

It should also be noted that most videotapes include voice or other audio information on a separate channel. Such information may be digitized and stored in personal computer 101 or in any other available memory. In the preferred embodiment of the invention, the audiosignal is received by computer 101 and is stored in one of two buffers by direct memory access. The first buffer receives one second worth of audio information and upon being filled with audio information, the buffer simultaneously time stamps the information in the buffer and sends out a signal to activate the second buffer to begin storing the next one second worth of information. A time stamp is an indication of the location of the videotape where the sound bit was recorded from. The time stamp is coordinated with the recordation of video in order to match the two elements. Once the second buffer is activated (undergoing DMA), the previously filled buffer is output to disk.

The stored audio data can be utilized during

playback, and can even be utilized to automatically annotate the stored frames. For example, each time a scene changes the computer can automatically store the first frame of the new scene and annotate it by including the few seconds of sound from before and after the scene change.

A convenient technique of storing and logging the digitally recorded video information in personal computer 101 is to do so in the form of at least three separate and distinct files. A first file is organized to store the actual reduced postage stamp images, where each record in this file comprises the pixels needed to make up the postage stamp. A second file, which may be thought of as a set of pointers, stores information sufficient to identify which postage stamp represents the first frame after a scene change and which postage stamps are not associated with scene changes. This file can be in simplistic form and may be nothing more than a list of numbers which indicate the record numbers in the postage stamp file representing frames recorded immediately after a scene change. This file would be used in the embodiment where only a subset of the recorded frames occur at scene changes. The embodiment and use of the file are described hereinafter.

A third file, denoted an annotation file, provides further convenience. The annotation file will be described below.

After the selected postage stamps are recorded, the display may display numerous postage stamps at a time. However, by the execution of a preprogrammed instruction, the display may go to one postage stamp and enlarge same to a full screen stamp. The operator can sequentially scroll through the postage stamps so that the sequence of different postage stamp size images are displayed on the monitor. A simple soft-

ware program allows the operator to select which frame is to be annotated. Annotation is a process whereby a message is typed onto the screen that explains the frame being viewed. For example, screen 504 of FIG. 5 may be annotated to state "Mr. Jones travels by plane to the big city".

One technique for designating the screen to be annotated, is to set up the software so that the middle frame is always selected for annotation. Although this is preferable, any screen can be set as the annotation screen, or the software can be programmed to vary same. Specifically, referring to FIG. 5, the operator scrolls through until the desired frame to be annotated is in the center of the monitor -- airplane 504 in FIG. 5. The operator presses a predetermined key on keyboard 209 in FIG. 2, which would then allow input of several lines of text at the bottom of the computer screen. A second predetermined key is pressed in order to exit annotation mode. If a different frame is desired to be annotated, the operator uses two predetermined keys to move the set of frames displayed so that the desired frame to be annotated is in the center of the screen. Alternatively, the desired frame can be selected with a mouse, or with any other technique.

FIG. 6 illustrates display screen that would appear on a monitor 103 of FIG. 1 in accordance with a first preferred embodiment. In this embodiment, the monitor, generally indicated at 601 includes a housing 602 and a display screen 603. Display screen 603 may take the form of a cathode ray tube (CRT) or a liquid crystal display (LCD). We will now discuss the display that a user would see on the screen 603. Each postage stamp 605 a-i depicts a separate frame received from a video signal. The frames each include a time stamp 604 which is a real time indication of

placement within the video signal. This stamp is used to log video tape, and is also helpful in finding cut points in the video tape. As shown herein, the middle screen, 605h is the screen that is annotated. The annotation box is indicated in the middle of the screen on the bottom, and is indicated as reference numeral 607. The annotation for screen 605e is "check pattern". Program indicators and help functions are also indicated on the bottom of the screen.

After annotation is complete, a separate file may be formed with all the annotated text and appropriate "pointers" (for example the time stamp) are provided to link the annotations to the proper frames with which they are associated. Thus, a subset of the stored frame will have annotations associated therewith. As an annotated frame appears in the center of the screen; its associated annotation will appear at the bottom of the screen as shown in FIG. 6. Predetermined keys can then be set up, via software, so that the displayed frames will move from one annotation to the next. For example, assume 1,000 frames are recorded, a selected 75 of which are annotated. The software can be programmed to display only the annotated frames. Accordingly, as the scroll keys are utilized, different frames are displayed, but those that are not annotated are skipped. Thus, the operator sees only a set of annotated frames, from which storyboards can be printed if desired.

FIG. 5 shows a front view of the monitor 101 displaying twenty five "postage stamp" sized frames. Three of the frames 501-503 are labelled. Each of the frames is a compressed version from one of the different scenes on the videotape. As can be seen from FIG. 5, the operator can quite easily and conveniently view a frame from each of a plurality of consecutive scenes.

In another enhancement, frames are stored at predetermined intervals and at scene changes. Thus, only a subset of the stored frames represent scene changes. The stored frames are then displayed on a computer screen, but only the frames which occur immediately subsequent to a scene change would be marked directly, for example, by placing a colored boarder around such frames. Thus, the computer screen would display a plurality of postage stamp size images, where several of the postage stamps would be "marked" as being scene changes with a colored border.

Returning to FIG. 2, view exploder 206 is utilized in order to provide a more sensitive scene change detector. When in the on position, view exploder 206 allows the sensitivity of the scene change detector to be increased. More particularly, rather than viewing the entire image to detect scene changes, view exploder 206 expands the picture so that a specified block of the picture is examined by scene change detector 203.

As shown in FIG. 3, the image 301 is comprised of a plurality of pixels. It is to be understood that only a small number of pixels are shown and that an actual video image is comprised of many more pixels. A particular block of pixels 302 is selected by view exploder 206. The image transmitted from view exploder 206 to scene change detector 203 is not the entire image, but rather an exploded view of the portion of the image made up by pixels 302.

For example, assume pixels 302 comprise 1/9th of the entire image. View exploder 206 would construct a 3x3 matrix of pixels for each single pixel in pixel group 302. The resulting image would then be a nine times expanded version of the portion of the image made up by pixel group 302. Put another way, the portion of the image made up by pixel group 302 would

be expanded to the size of the entire image.

As scene change detector 203 examines the incoming video signal for scene changes, scene change detector 203 will, in actuality, only be examining the diminished portion of the image comprised of pixel group 302. Therefore, any change in this portion will result in a scene change being detected, even though the remainder of the entire image does not change. Therefore, scene changes can be detected which would not ordinarily be detected without the view exploder 206.

For example, a videotape is often made from surveillance data recorded by a video camera. The camera surveys a wide area, but if an intruder enters only a small portion of the area covered by the camera, the scene will not necessarily change. As an example, a door may open, but the door is only a very small percentage of the image of the entire room. A normal scene change detector may not recognize this as a scene change because most of the image remains the same. With view exploder 206 in use, the particular portion containing the door could be blown-up so that there would be a big difference between an image comprised mostly of a closed door, and that comprised mostly of an open door. Hence, the scene change could be detected.

It should be noted that rather than view exploder 206, different techniques can be utilized to cause scene change detector 203 to examine only a portion of the scene. For example, a counter could be utilized which counts pixels, and scene change detector 203 would be arranged to only examine certain pixels, namely those in pixel group 302.

FIG. 4 depicts a block diagram of a scene change detector which may be utilized in accordance with the present invention. It should be noted that the par-

ticular scene change detector is not critical to the operation of the invention and many other scene change detectors may be utilized.

In operation, the video signal 207 is fed in parallel through low pass filter 400 and sync stripper 418. The sync stripper is a straightforward signal processor which removes the synchronization signal that is present in all National Television Standards Committee (NTSC) signals. Low pass filter 400 removes any noise and/or spikes in the video input signal. In the preferred embodiment, the video input signal is filtered down to 100 kilohertz. The low pass filtered signal is then transferred to analog-to-digital (A/D) converter 402, which samples the signal at a rate in excess of the required Nyquist rate, as is known in all digital signal processing systems. In a typical embodiment, it has been found that a six bit sample approximately every eight microseconds will suffice.

The digital samples are sent to a six-bit memory with 2,048 locations as indicated by memory unit 406 in FIG. 4. The illustrative memory can hold 2,048 samples, although other size memories can certainly be used. Approximately 1,500 to 1,700 samples are stored from each video frame and the samples stored are corresponding samples from consecutive frames.

Subtractor 408 is arranged to receive its inputs from the present sample, which is output by A/D converter 402, as well as from the corresponding sample from the previous frame which is conveyed from memory 406 to the other input of subtractor 408. The difference between samples represents the amount of change that a particular incrementally small portion of the image has undergone from one frame to the next. The output of the subtractor (incremental change is the sample) is delivered to absolute value generator 410, which outputs the positive value corresponding to the

foregoing difference, whether positive or negative, between the corresponding samples of consecutive frames.

Adder 412, initially set to zero, begins summing the positive values of the differences output by absolute value generator 410. The first difference received by adder 412 is added to zero and the resulting value (the first difference itself) is placed into latch 414. Latch 414 may simply be a memory location in the random access memory of a microprocessor-controlled system.

During the next sampling period, absolute value generator 410 transmits the absolute value of the next difference to adder 412. Adder 412 then combines this second absolute value with the first one, which is fed back via path 418, and adder 412 then places into latch 414 the sum of the absolute value of the two differences. The process repeats for all 1,500 to 1,700 samples taken for a particular frame. Each time adder 412 adds, it is adding the absolute value of the difference between the present sample and the corresponding sample from the previous frame, to the sum of all previous absolute values associated with a particular video frame being sampled. It can readily be appreciated that at the end of sampling of an entire video frame, latch 414 will hold the sum of the absolute values of all the differences between each sample for the particular frame and the corresponding sample for the previous frame.

The value stored in latch 414 at the end of an entire frame therefore represents an amount of change which has occurred between the previous frame and the present frame. However, it is only a representative sample since not every bit is sampled. The value obtained can be thought of as a representation of the derivative of the image. This quantity is transferred

to field latch 416 and thereafter transmitted to the host computer.

The host computer is programmed to determine that a scene change has occurred if the value transmitted from field latch 416 is greater than a predetermined value. After such determination and at the end of each field, adder 412, latch 414, and field latch 416, are reset to zero for the new frame.

The predetermined value can be raised or lowered according to the type of video signal received. For example, if the video is a monologue or an interview, the predetermined value is set relatively low -- in order to be very sensitive. Alternatively, if the video is an action piece with explosions, car chases, etc. then the predetermined value is set higher -- in order to have less sensitivity.

Timing and control circuit 420 is utilized to ensure that sync stripper 418 is activated when the synchronization signal is present so that the proper part of the video input signal is stripped rather than the usable information therein. Hardware for stripping the sync signal is readily available and well known in the television art.

Address generator 422 controls which of the 2,048 samples stored in memory 406 is input to subtractor 408. This is a straightforward task which may be accomplished with a simple algorithm programmed into a basic microprocessor. The address generator ensures that for each sample that emerges from A/D convertor 402, the corresponding sample from the previous frame is read out of memory 406.

It should be noted that when the video picture is unchanging, or changing rather slowly, the output from field latch 416 will be close to zero for successive frames. When a pan or a zoom is occurring in the video image, a step change will occur at the output of

field latch 416 because successive frames which initially exhibit little change will suddenly initiate a constant change between successive frames for several frames in a row. Finally, a scene change shows up as a spike in the output of field latch 416. The spike can be detected by the host computer and processed in order to decide whether or not there has been a scene change, in accordance with the predetermined threshold set in the computer.

In other words, the cut detect algorithm works by reviewing five consecutive values from field latch 416. If the middle value exceeds each of the other four values by an operator-determined cut threshold, then a cut is flagged as having taken place at the frame corresponding to the middle value of the five frames. This threshold value can be varied to the sensitivity desired by the user.

Fades, wipes, tilts, pans, zooms or movement of objects or persons in the video field are all characterized by frame-to-frame activity that continues over several frames. This appears as relatively constant activity in the change values (output of field latch 416). However, no spikes would appear, therefor, the cut algorithm filters this out.

By way of example, in the preferred embodiment, a camera shot of an empty room produces a stream of numbers from field latch 416 of between 400 to 500. As an actor walks on-screen, there is a step change in value at field latch 416 to between 1500 to 2000. The cut detect algorithm would filter this out. However, if there is a cut close-up of the actor, there would be a spike in the 20,000 range, which would be detected by the algorithm.

Whip pans, where the camera is rapidly swung from one point-of-view to another, result in step changes of unusually high value. The whip pan detect algo-

rithm is design to register a whip pan when a predetermined level of activity (output of field latch 416) is detected from a predetermined number of consecutive fields.

In another embodiment, video frames are stored periodically. For example, every $n^{th}$ frame may be stored, where n is a number greater than 1. The stored frames may then be simultaneously displayed on the video monitor or computer, thereby providing a smooth visual progression of scenes. The scene change detector is only utilized to store and display scene change information.

In order to log a large number of scenes into a computer which contains a finite amount of storage capacity, the signal received must be compressed. Reference is now made to FIG. 7, wherein a full bandwidth NTSC video signal 701 which includes approximately 14 megabytes per second throughput for transmission through a digital system, is provided. Obviously, any system capable of recording 14 megabytes of data per second for an extended period of time (for example 3 hours per movie) would be considered a super computer and would cost hundreds of thousands of dollars. Prior art compression techniques can reduce signals by ratios of as much as 200:1. However, adverse effects of compression are normally noticed at approximately 4:1 and become objectionable at approximately 12:1.

The present invention undergoes what is known in the art as "destructive compression" inasmuch as the stored image lacks color, clarity and continuity. However, the ultimate compression is on the magnitude of 1000:1 and provides a suitable image for its purpose. This is how the present invention is capable of logging multiple hours worth of videotape with a personal computer, having as little as an INTEL 386 or

486 microprocessor.

Video signal 701 passes through a low pass filter 703 to produce a low pass filtered video signal. The luminance bandwidth of a NTSC video is on the magnitude of 4.2 megahertz. The low pass filter reduces the luminance bandwidth to 1.2 megahertz. The filtration removes the color information from the video signal and produces a picture in many shades of grey.

The signal is next sampled 705. The sampled lines undergo 2:1 decimation 707 vertically. This results in 120 by 184 six-bit monochrome pixels which represent one image. Prior to storing this information, each group of four six-bit pieces of data are compressed 709 to form three eight-bit bytes. This result is then stored 711 -- on computer disk or the like. The result is that video frames are stored at a rate of 15,650 bytes per second as opposed to 14.3 megabytes for one second (60 fields) of full-color, full-bandwidth television.

With particular reference to FIG. 8, another embodiment of the invention depicts monitor 801 displaying six frames of video 803 a-f in the middle portion of the video screen. This embodiment includes audio waveform 805 illustrated on the display screen. The audio waveform allows the user to place a cut marker at the exact point desired with respect to the audio information. As stated above, this embodiment is most useful in the case of an interview, where there are very few, and possibly no scene changes detectable by the video scene change detector. For example, the user may indicate a scene change at the end of each completed question and answer. Furthermore, annotation block 807 can be completed to indicate the subject matter of the question and answer.

The audio signal uses 184 dots to represent one second's worth of audio. Each second of audio is

typically 8,000 eight-bit samples, so each dot on the screen needs to represent 67 audio samples. The software scans the dots and determines the maximum and minimum value during the period and draws a vertical line to connect the dots. This line is normalized to fit in the appropriate window on the screen. This process is continually reported for all the audio information to obtain a continuous graph.

When the audio sound track is quiet, the line created will be short and close to zero. Alternatively, when the audio sound track is loud, the minimum and maximum values will be large, so the video graph of the sound track will be much thicker.

Accordingly, when an operator is logging a tape that includes, for example interviews or music video which may be edited in large part by the audio component change, the present system provides a great advantage. The operator uses the computer display and the mouse to walk through the video graph of the sound track to quickly and accurately determine the exact frame required to achieve a desired result.

In accordance with the teachings described herein, a video scene detection and logging system could be constructed which operates to effectuate recording and logging of the scenes by the operator. As the videotape player 102 is played, the scenes change periodically. As the scenes change at least one frame from each scene is captured and compressed, and stored by storage unit 201. After the entire videotape is played, the plurality of scenes may be displayed in groups of, for example, 25, 9 or 6 at a time on a single screen as shown in FIGS. 5, 6 and 8, respectively.

The operator can then view the entire screen and get a basic feel for the flow of images on the videotape since he has at least one frame from each scene

in front of him. The device may also be configured to print out an edit list that could be an input to an editing system. Additionally, in any of the embodiments, the displayed frames may include information such as frame number, time, or any other identifying information. The operator can then determine which frames to edit out and/or if the order of the scene should be rearranged, this can be done using typical editing and splicing techniques on any one of many commercially available machines.

It is understood that while the above describes the preferred embodiments of the invention, various other modifications and/or additions may be made without violating the spirit and scope thereof. For example, different types of video input signals may be utilized, different designs for the required software algorithms may be employed, etc. Additionally, combinations of the embodiments may be used, such as storing every $n^{th}$ frame and additionally storing a frame when the scene changes. Since certain changes may be made in carrying out the above invention without departing from the spirit and scope of the invention, it is intended that all matter contained in the above description shall be interpreted as illustrative and not in a limiting sense.

It is also to be understood that the following claims are intended to cover all of the generic and specific features of the invention herein described and all statements of the scope of the invention which, as a matter of language, might be said to fall therebetween.

## CLAIMS

What is claimed is:

1. A video scene change detection and logging apparatus for logging information from a video source comprising:

means for detecting scene changes in a video signal received from said video source;

means for recording selected frames from a plurality of scenes from said video source;

means for displaying a reduced version of each of said recorded frames on a video monitor thereby providing convenient viewing by an operator;

means for linking each of said recorded frames to a location on said video source; and

means for compiling a list of the information included on the video source.

2. The apparatus of Claim 1, wherein said means for detecting includes means for comparing successive frames in said video signal.

3. The apparatus of Claim 2, wherein said means for comparing includes means for comparing only a portion of each frame with a corresponding portion of a previous frame.

4. A video scene change detection and tape logging apparatus for logging information from a video source, wherein said information includes a plurality of video frames and each video frame includes a plurality of frame portions, comprising:

means for selectively recording predetermined frames received from said video source;

means for selectively isolating predetermined portions of said recorded frames;

means for comparing a first selectively isolated frame portion with a corresponding second selectively isolated frame portion of an adjacent video frame; and

means responsive to said comparing means for determining that a scene change has occurred.

5. A computerized video scene change detection and logging apparatus for logging video information from a video information source comprising:

means for converting said video information into digital information;

means for recording every $n^{th}$ frame of video information from a video information source in digital form where $n > 1$; and

means for simultaneously displaying a plurality of said recorded frames on a video monitor, such that the video information is compressed since all frames between every $n^{th}$ frames are removed and so that said recorded information can be digitally accessed from said recording means at a rapid rate.

6. A method of detecting scene changes in a video source, and of logging information from the videosource, said method comprising the steps of:

comparing at least a portion of successive scene changes in a video signal;

recording selective frames from a plurality of scenes;

displaying a reduced version of each of said recorded selective frames on a video monitor, thereby providing convenient viewing by an operator;

linking each of said recorded selective frames to a location on the video source; and

compiling the recorded information.

7. A method of detecting video scene changes and logging video information from a video source, wherein said information includes a plurality of video frames and each video frame includes a plurality of frame portions, comprising the steps of:

selectively recording predetermined frames from said video source;

selectively isolating predetermined portions of said recorded frames;

comparing a first selectively isolated frame portion with a corresponding second selectively isolated frame portion; and

determining that a scene change has occurred when said step of comparing indicates there is at least a predetermined difference between said first selectively isolated frame portion and said second selectively isolated frame portion.

8. An apparatus for logging audio and video information from an audio/video source, comprising:

means for recording selected video information received from said audio/video source;

means for recording said audio information received from said audio/video source;

means for displaying said selected video information, to be viewed by an operator; and

means for automatically matching said audio information to associated video information, so that at least one of said audio information and said video information may be logged.

9. The apparatus of claim 8, further including means for generating a video signal corresponding to said audio information to be displayed on said display means.

10. The apparatus of claim 8, wherein said displaying means displays a plurality of frames of said video information simultaneously.

11. The apparatus of claim 8, further including a first storage means for storing information representative of individual video frames, a second storage means for storing information sufficient to identify which of said individual stored frames represent scene change information, and a third storage

means for storing information to identify the subject matter of the stored frame.

12. An apparatus for logging audio and video information from an audio/video source comprising:

means for detecting scene changes in said video information received from said audio/video source;

means for recording selected frames of video from a plurality of scenes from said audio/video source;

means for recording said audio information from said audio/video source; and

means for linking said audio and video information, so that at least one of said audio information and said video information may be logged.

13. An apparatus for logging audio and video information from an audio and video information source comprising:

means for receiving selected video information from said audio/video source;

means for receiving selected audio information from said audio/video source;

means for generating a video signal corresponding to audio information; and

means for displaying said video signal corresponding to said audio information and simultaneously displaying a plurality of selected frames of said video information correspond to the audio information.

14. The apparatus of claim 13, further including means for playing back the recorded audio information indicating the video frame most closely associated with the audio information.

15. A logging device for logging audio information from an audio source comprising:

means for recording audio information from

an audio source;

means for generating a video signal representative of the audio information, wherein the video signal distinguishes between different volumes of the audio information;

means for indicating positions on said video signal that are associated with particular audio information from the audio source; and

means for generating a list of said positions.

16. A method of logging video tape, comprising the steps of:

receiving audio information and video information from a videotape;

recording said audio and video information in digital form;

generating a video signal representative of the audio information;

marking the video signal to indicate cut points of the videotape; and

compiling a list of said cut points.

17. A method of compressing a video frequency signal having a high frequency component and a low frequency component comprising the steps of: filtering said signal to remove said high frequency component;

sampling said filtered video signal;

decimating said sampled video signal at essentially a ratio of 2:1; and

compressing said decimated signal.

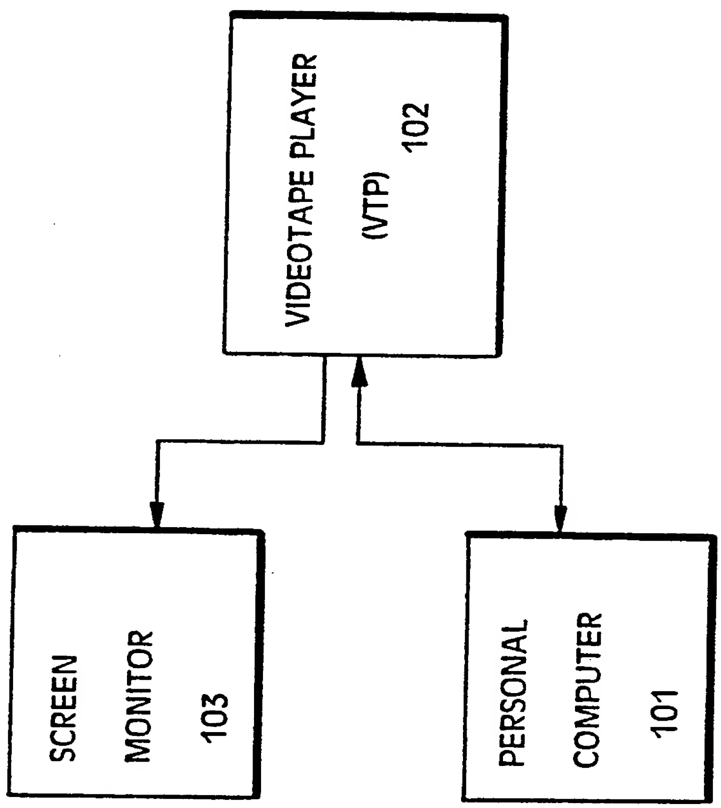

VIDEOTAPE PLAYER
(VTP)
102
SCREEN
MONITOR
103
PERSONAL
COMPUTER
101

FIG.1

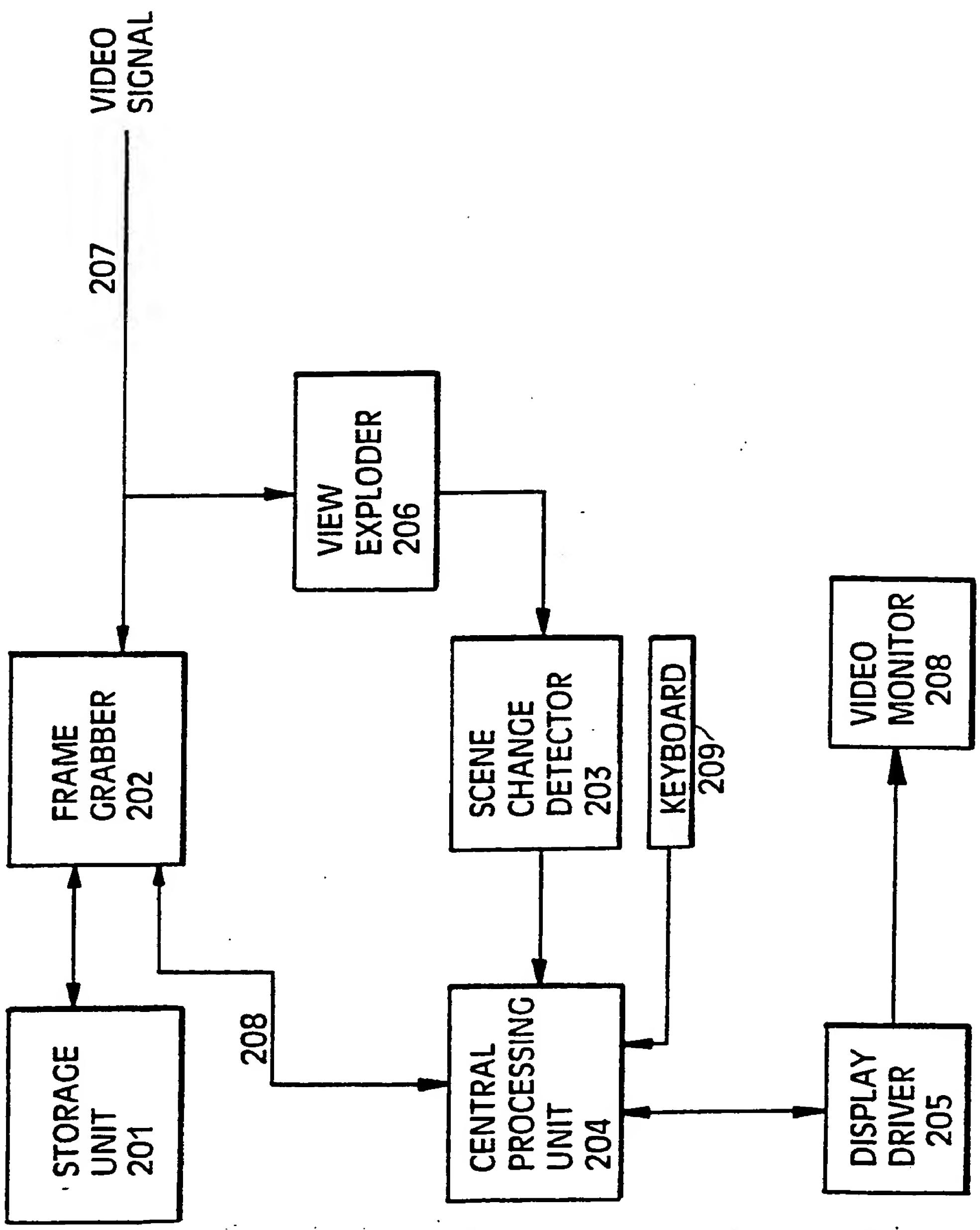

VIDEO SIGNAL
207
VIEW EXPLODER 206
FRAME GRABBER 202
SCENE CHANGE DETECTOR 203
KEYBOARD 209
VIDEO MONITOR 208
STORAGE UNIT 201
208
CENTRAL PROCESSING UNIT 204
DISPLAY DRIVER 205

FIG. 2

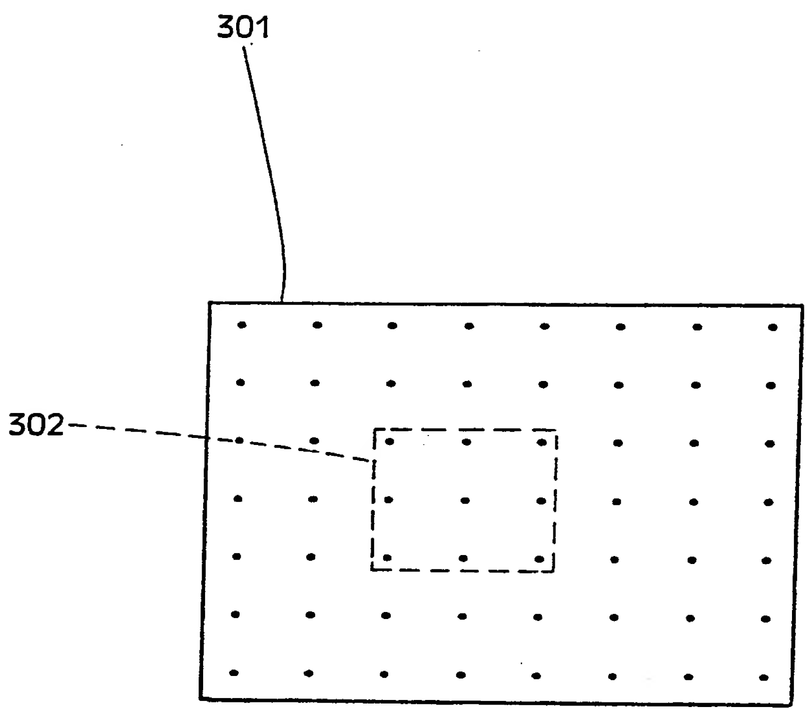



FIG. 3

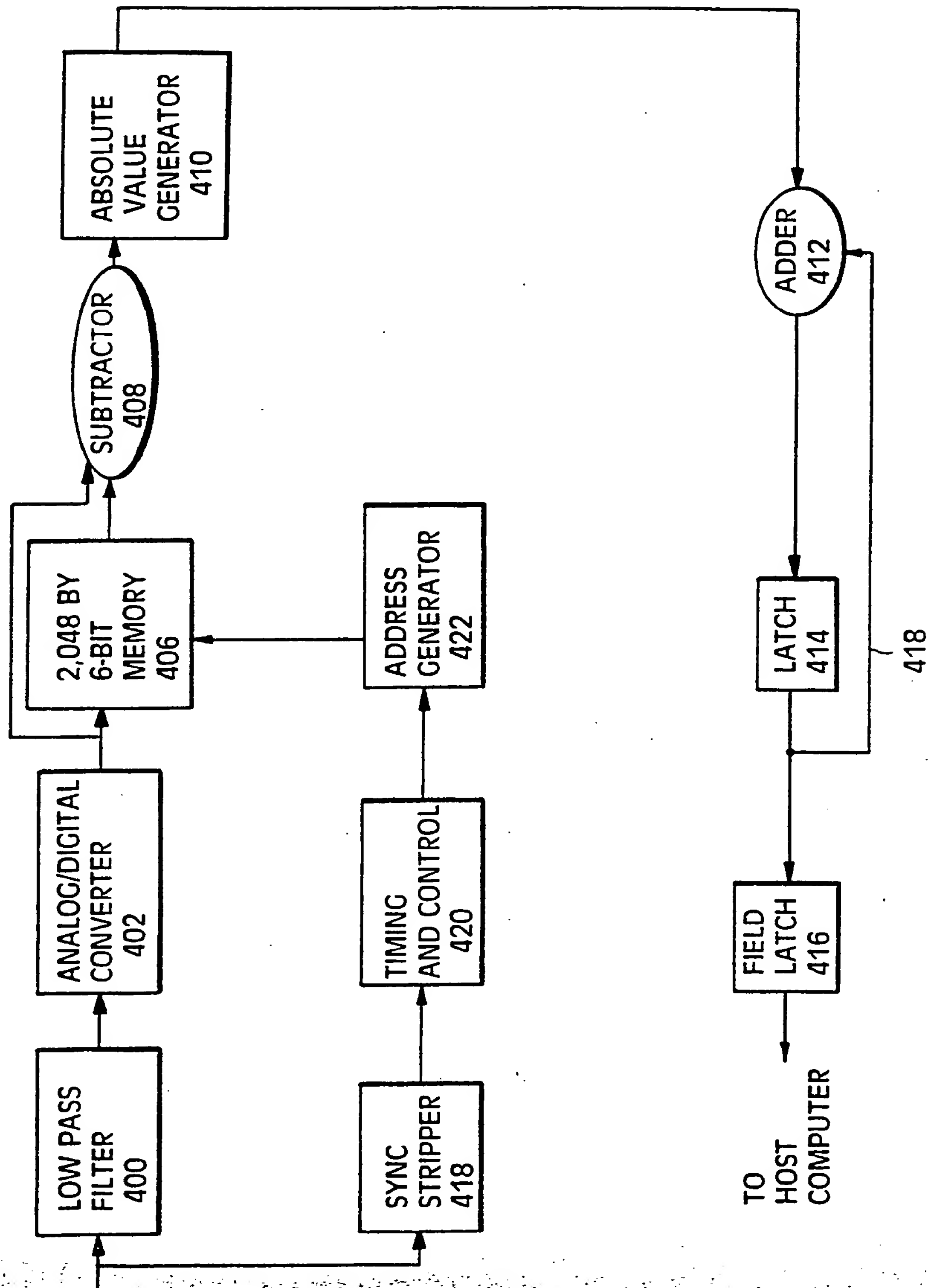

LOW PASS FILTER 400
ANALOG/DIGITAL CONVERTER 402
2,048 BY 6-BIT MEMORY 406
SUBTRACTOR 408
ABSOLUTE VALUE GENERATOR 410
SYNC STRIPPER 418
TIMING AND CONTROL 420
ADDRESS GENERATOR 422
ADDER 412
LATCH 414
418
FIELD LATCH 416
TO HOST COMPUTER

FIG. 4

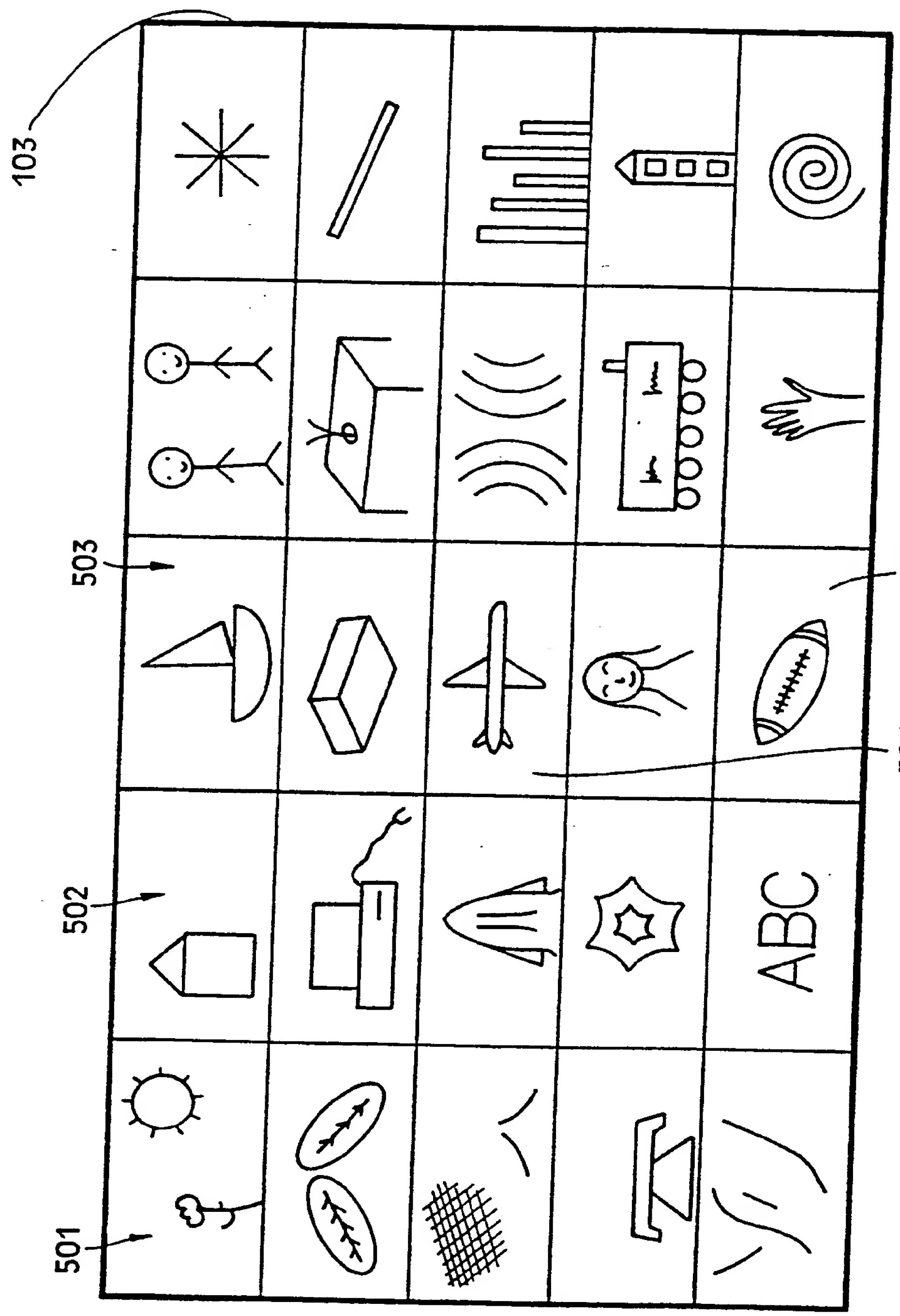



FIG. 5

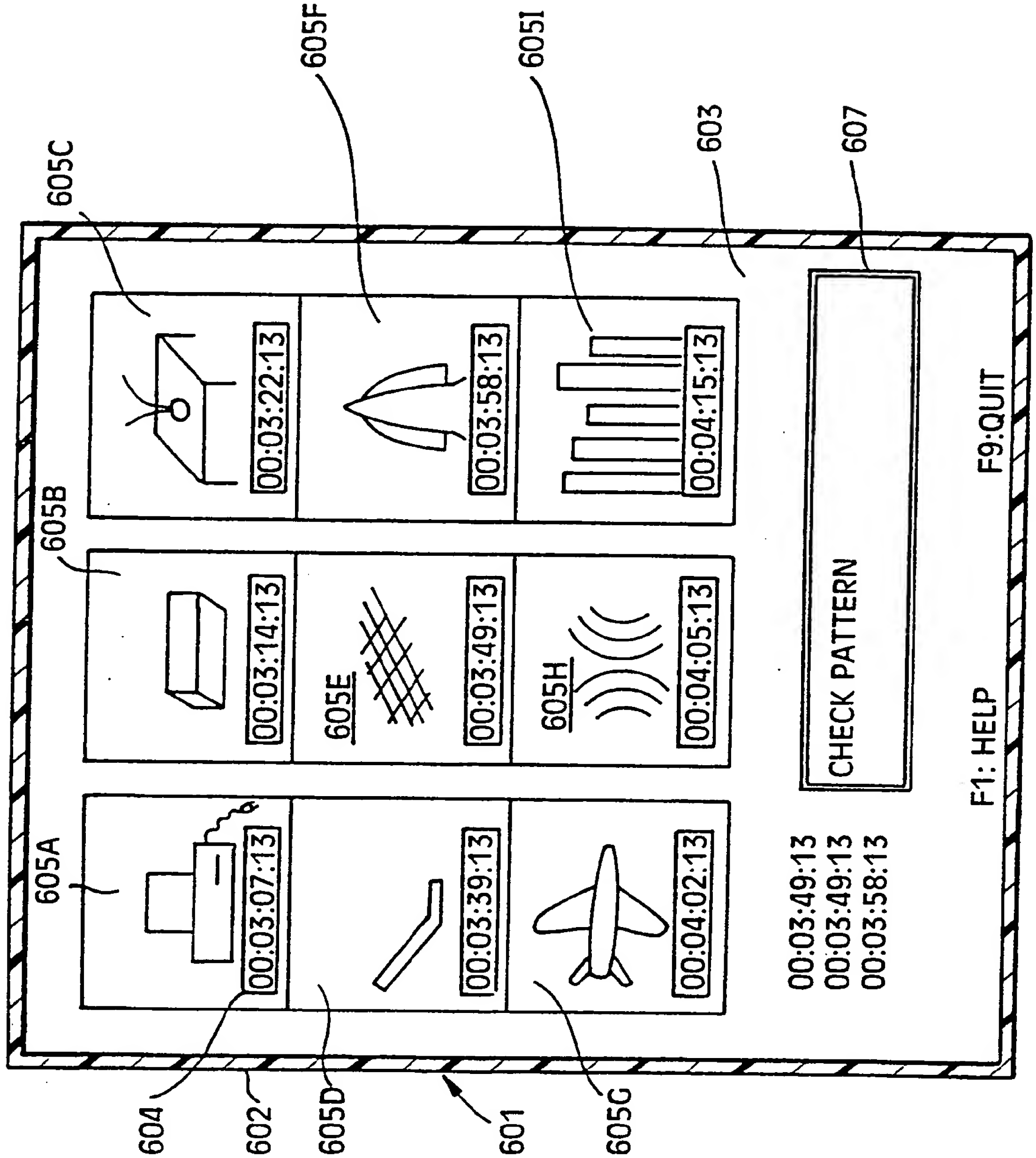

605C
605F
605I
603
607
605B
605A
00:03:07:13
00:03:14:13
00:03:22:13
00:03:39:13
605E
00:03:49:13
00:03:58:13
00:04:02:13
605H
00:04:05:13
00:04:15:13
00:03:49:13
00:03:49:13
00:03:58:13
CHECK PATTERN
F1: HELP
F9:QUIT
604
602
605D
601
605G

FIG. 6

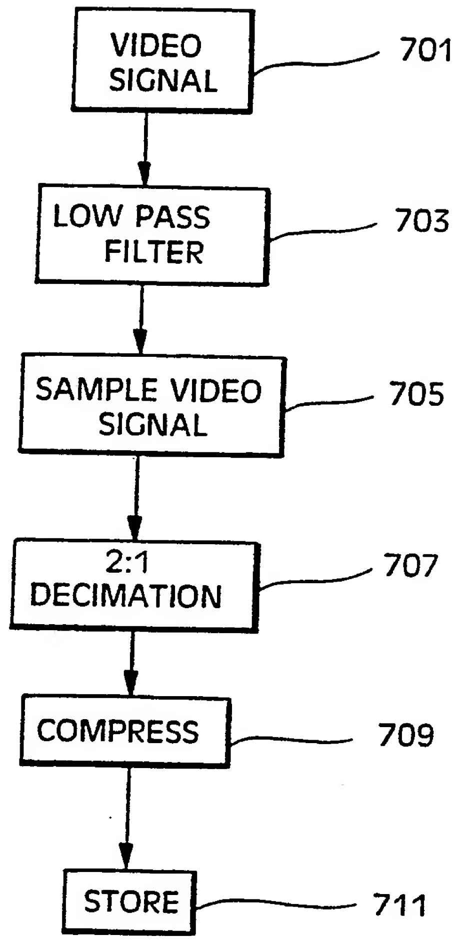

VIDEO SIGNAL
701
LOW PASS FILTER
703
SAMPLE VIDEO SIGNAL
705
2:1 DECIMATION
707
COMPRESS
709
STORE
711

FIG. 7

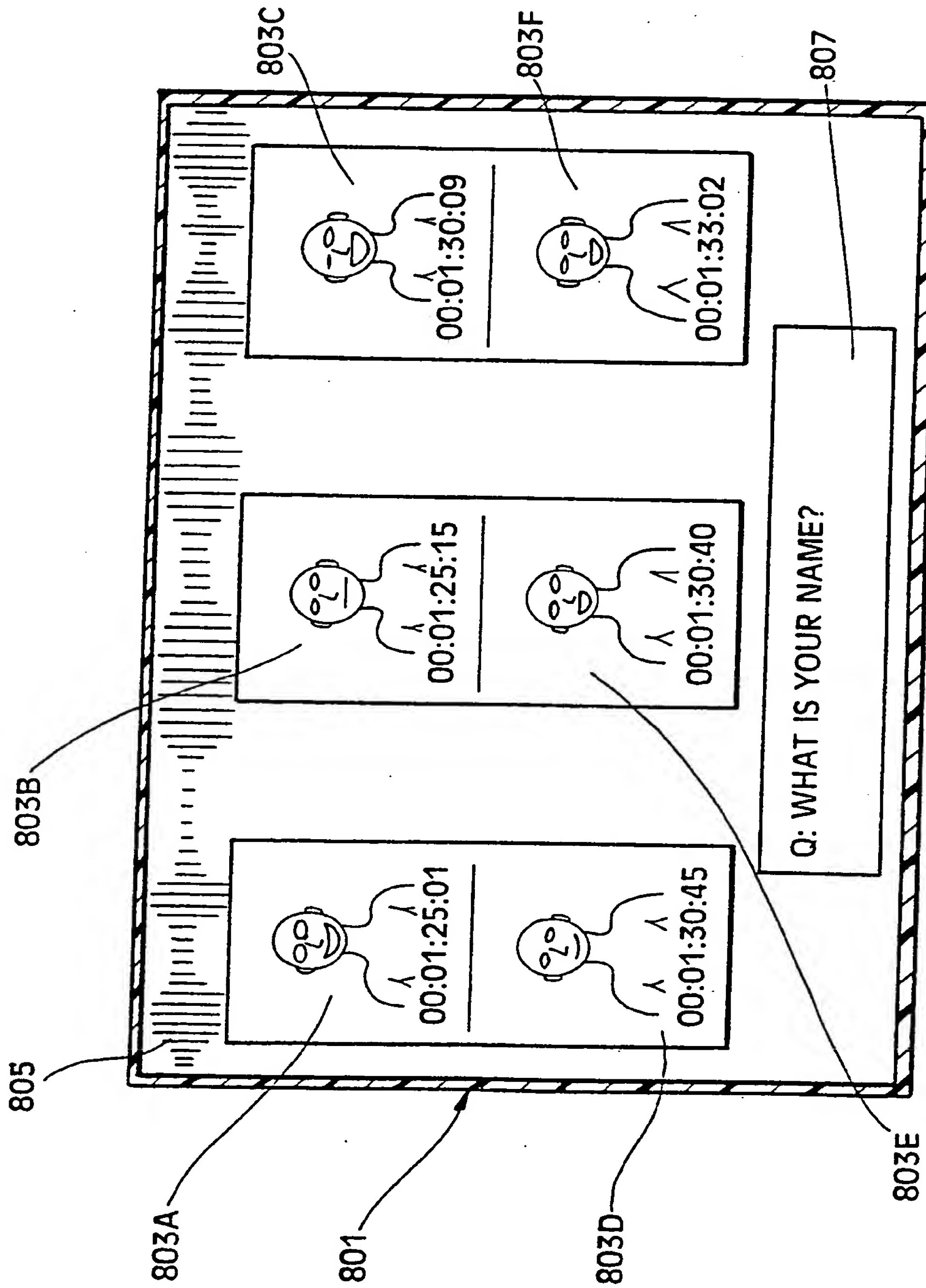



FIG.8



SUBSTITUTE SHEET

# INTERNATIONAL SEARCH REPORT

International application No.
PCT/US93/11267

**A. CLASSIFICATION OF SUBJECT MATTER**

IPC(5) :H04N 9/79, 5/76
US CL :358/311, 335; 360/13
According to International Patent Classification (IPC) or to both national classification and IPC

**B. FIELDS SEARCHED**

Minimum documentation searched (classification system followed by classification symbols)

U.S. : 358/311, 335, 906, 909, 182, 183, 22; 360/13, 14.1,33.1, 35.1

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practicable, search terms used)

**C. DOCUMENTS CONSIDERED TO BE RELEVANT**

| Category* | Citation of document, with indication, where appropriate, of the relevant passages | Relevant to claim No. |
|---|---|---|
| X<br>----<br>Y | US, A, 4,920,423 (Shiota) 24 April 1990, Figure 4 and Figure 1, components 16, 7 and 25, col. 3, lines 16-25 and 55-62, col. 6, lines 34-48. | 1-2, 5<br>----------<br>3-4, 6-16 |
| Y, P | US, A, 5,233,438 (Funahashi et al) 03 August 1993, col. 1, lines 1-51. | 8-16 |
| Y | US, A, 4,864,396 (Martens) 05 September 1989, col. 3, lines 54-68, col. 4, lines 1-62. | 17 |

☐ Further documents are listed in the continuation of Box C. ☐ See patent family annex.

* Special categories of cited documents:
"A" document defining the general state of the art which is not considered to be part of particular relevance
"E" earlier document published on or after the international filing date
"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)
"O" document referring to an oral disclosure, use, exhibition or other means
"P" document published prior to the international filing date but later than the priority date claimed
"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention
"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone
"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art
"&" document member of the same patent family

| | |
|---|---|
| Date of the actual completion of the international search<br>07 January 1994 | Date of mailing of the international search report<br>MAR 15 1994 |
| Name and mailing address of the ISA/US<br>Commissioner of Patents and Trademarks<br>Box PCT<br>Washington, D.C. 20231<br>Facsimile No. NOT APPLICABLE | Authorized officer<br>TOMMY CHIN<br>Telephone No. (703) 305-4780 |